We Claim:

1. A method for identifying a protein through amino acid sequences of one or more query peptides generated from the protein comprising:
   (a) translating amino acid sequences of one or more query peptides to all possible codons from which the peptides can be synthesized to prepare strings of codons;
   (b) searching known nucleic acid sequences to locate one or more known nucleic acids that comprise regions that match the strings of codons; and
   (c) ranking two or more matching nucleic acids to identify nucleic acids that are true coding regions for the protein to thereby identify the protein.
2. A method of claim 1 wherein the amino acid sequences of the query peptides are obtained from the spectra produced by mass spectrometry of the peptides.
3. A method of claim 1 or 2 wherein in (b) the searching comprises simultaneously providing the strings of codons as parallel queries to a database of known nucleic acid sequences.
4. A method of claim 1, 2, or 3 wherein in (b) the searching further comprises locating one or more known nucleic acids that comprise regions that match reverse complements of the strings of codons.
5. A method of any preceding claim wherein in (c) the ranking is based on a comparison of masses of peptides translated from sequences in proximity to the regions in the known nucleic acids that match the strings of codons with masses of peptides of the protein other than the query peptides.
6. A method of any preceding claim wherein the strings of codons comprise wildcards.
7. A method of any preceding claim wherein in (c) the ranking comprises the following steps:
   (a) calculating the masses of peptides translated from sequences in proximity to the regions in the known nucleic acids that match the strings of codons;
   (b) comparing the masses calculated in (a) with masses of peptides of the protein other than the query peptides, or fragments thereof, to identify peptides with matching masses;
   (c) assigning scores to each matching mass and accumulating the scores for all matching masses in proximity to the regions in the known nucleic acids that match strings of codons; and
   (d) ranking two or more nucleic acids that match the strings of codons based on the accumulated scores to identify potential nucleic acids encoding the protein to thereby identify the protein.
8. A method of claim 7 wherein in (b) the masses of peptides of the protein other than the query peptides or fragments thereof are identified through mass spectrometry.
9. A method of claim 8 wherein the masses of the peptides are identified in a precursor ion scan.
10. A computer implemented system for identifying a protein through amino acid sequences of one or more query peptides generated from the protein comprising:
   (a) a search engine for locating regions of known nucleic acid sequences that match strings of codons translated from one or more query peptides;

(b) a mass calculator for calculating masses of peptides translated from sequences in proximity to regions in known nucleic acid sequences that match the strings of codons;

(c) optionally a scoring unit for (i) comparing masses calculated in (b) with masses of peptides of the protein other than the query peptides to identify peptides with matching masses; (ii) assigning scores to peptides with matching masses; and (iii) accumulating scores for all matching masses in proximity to or around the regions located in (a) to evaluate the likelihood that a region is a true coding region for the protein.

11. A method for identifying a protein comprising:

(a) providing amino acid sequences of peptides generated by mass spectrometry of the peptides cleaved from the protein;

(b) translating amino acid sequences of one or more query peptides to all possible codons from which the peptides can be synthesized to prepare strings of codons;

(c) searching known nucleic acid sequences to locate one or more known nucleic acids that comprise regions that match the strings of codons; and

(d) optionally ranking two or more matching nucleic acids located in (c) by

(i) calculating the masses of peptides translated from sequences in proximity to regions in the known nucleic acids that match the strings of codons;

(ii) comparing the masses calculated in (i) with masses identified by mass spectrometry for peptides of the protein other than the query peptides to identify peptides with matching masses;

(iii) assigning scores to each matching mass and accumulating the scores for all matching masses in proximity to regions in known nucleic acids that match the strings of codons; and

(iv) ranking two or more known nucleic acids that match the strings of codons based on the accumulated scores to identify potential nucleic acids encoding the protein to thereby identify the protein.

12. A method of any preceding claim wherein the query peptides are tryptic peptides.

13. A programmable hardware employing a method as claimed in any preceding claim.

14. A hardware acceleration system for identification of a protein comprising a generic circuit board capable of being plugged into a computing device wherein the circuit board comprises logic chips and memory wherein the memory comprises nucleic acid sequence information, and the chips provide means to search through the nucleic acid sequence information for regions matching strings of codons translated from one or more query peptides provided as input to the computing device.

15. A hardware acceleration system for identification of a protein comprising a generic circuit board capable of being plugged into a computing device wherein the circuit board comprises logic chips and memory wherein the memory comprises nucleic acid sequence information, and the chips provide means to search through the nucleic acid sequence information for patterns matching a query that has been provided to the computing device as input from a mass spectrometer.

16. A method as claimed in any preceding claim implemented using field programmable gate array (FPGA) technology.

17. A method as claimed in any preceding claim implemented using application-specific integrated circuit (ASIC) technology.

18. A method as claimed in any preceding claim wherein known nucleic acid sequences comprise a genome, in particular human genome.

19. A database comprising a set of masses corresponding to the masses of the query peptides and the peptides translated from a matching region in proximity to or around a known nucleic acid generated in accordance with a method of any preceding claim.

20. Computerized representations of information generated using a method of any preceding claim, including any electronic, magnetic, or electromagnetic storage forms of the information needed to define it such that the information will be computer readable for purposes of display and/or manipulation.

21. A computer comprising a machine-readable data storage medium comprising a data storage material encoded with machine readable data wherein said data comprises information generated using a method of any preceding claim.

22. A method for presenting information pertaining to nucleic acids that potentially encode a protein the method comprising the steps of:

(a) providing an interface for entering query information generated from mass spectrometry relating to amino acid sequences of peptides generated or cleaved from the protein;

(b) examining records in a database of known nucleic acid sequences to locate regions in the nucleic acid sequences matching strings of codons translated from the entered query peptides' amino acid sequence information;

(c) displaying the data relating to the matched string of codons and regions in the nucleic acids; and

(d) optionally displaying the masses of the peptides generated from mass spectrometry and the masses of peptides encoding regions in proximity to the regions of known nucleic acids that match the string of codons.

23. A computer program product comprising a computer-usable medium having computer-readable program code embodied thereon for effecting the steps of a method of any preceding claim.

24. A method of using a method, system, programmable hardware, database, product, or computer as claimed in any preceding claim to identify proteins associated with disease or that can be used in drug design.

25. A method of using a method, system, programmable hardware, database, product, or computer as claimed in any preceding claim to identify proteins in samples from patients.